Identifying and attributing common data quality problems: temperature and precipitation observations in Bolivia and Peru

Stefan Hunziker *[1]*, **Stefanie Gubler** *[2]*, **J. Calle** *[3]*, **I. Moreno** *[3]*, **M. Andrade** *[3]*, **F. Velarde** *[3]*, **Laura Ticona** *[3]*, **G. Carrasco** *[4]*, **Y. Castellón** *[4]*, **Clara Oria Rojas** *[5]*, **S. Brönnimann** *[1]*, **M. Croci-Maspoli** *[2]*, **T. Konzelmann** *[2]*, **Mario Rohrer** *[2]*

Contact author: stefan.hunziker@giub.unibe.ch

[1] Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland *[2]* Federal Office of Meteorology and Climatology MeteoSwiss, Zürich, Switzerland *[3]* Laboratorio de Física de la Atmósfera, Instituto de Investigaciones Físicas, Universidad Mayor de San Andrés, La Paz, Bolivia *[4]* Servicio Nacional de Meteorología e Hidrología de Bolivia, SENAMHI *[5]* Servicio Nacional de Meteorología e Hidrología del Perú, SENAMHI

b UNIVERSITÄT BERN

OESCHGER CENTRE CLIMATE CHANGE RESEARCH

1 Motivation

- In-situ climatological observations are essential for studies on climate trends and extreme events.
- Communities in the Central Andes are particularly vulnerable to extreme events and climate change [e.g. 1, 2].
- Reliable observation data is needed to investigate changes in climate and extremes.

2 Challenges

- Data from weather station in Bolivia and Peru is often affected by data quality problems.
- The station network in the Central Andes is sparce.
- Quality problems in such networks differ at least partly from the errors found in state of the art western observation networks due to different
  - instruments,
  - station maintenance,
  - observation practices,
  - observer instructions, and
  - data processing and transmission.
- Information on stations and station histories are fractional or missing.

3 Metadata

- Reliable metadata is essential for any further use of the station data.
- Metadata may improve the results of further data processing and analysis such as data homogenization [e.g. 3].
- If metadata are lacking or missing, **station visits** (Fig. 1) are useful to
  - assess the actual state of the station (Fig. 2),
  - reconstruct station history (e.g. by talking to the observer), and
  - detect the sources of data errors.

Fig. 1: Station visit in Chiripaca, Bolivia.

Fig. 2: Rain gauge Vino Tinto, Bolivia. The data from this station do not show any abnormality. However, the rain gauge is sited in a yard at the height of the roof. The measurements are likely to be affected by splash-in (roof) and complex wind fields (walls, roof). Due to these side specific influences, the data may not be suitable for a wide range of applications.

4 Common data quality problems

There are several common quality problems found in Central Andean station data, such as the reduction of climatological variability, rounding inconstancies, missing low precipitation values, weekly precipitation cycles (or precipitation accumulations), and several transcription errors. Two cases of errors will be presented here.

Usually, time series affected largely by errors and/or suspicious values are removed from further analyzes [e.g. 4, 5]. Because of the scarce data availability in the Central Andes, analyzing errors and potentially detecting the reason of occurrence may allow error correction and decisions on further use of the data for specific applications.

A) Missing temperature intervals

A relatively high number of stations (around 20% in the Peruvian study region) is affected by missing temperature intervals, mostly around 0°C. A special case of this error was found in minimum temperatures (TN) at the station Progreso, Peru. Since July 2003, there are no observations of temperature between -2 and 2°C (Fig. 3, above). A station visit in Progreso showed that the observer correctly reported the value at the right end of the rod if the balance point of the rod was above 0°C, and erroneously on the left if it was below 0°C (Fig. 3, below). Thanks to this finding, the error can be corrected by adding 4.2°C to each of the negative values, and the data can be used for further analyzes.

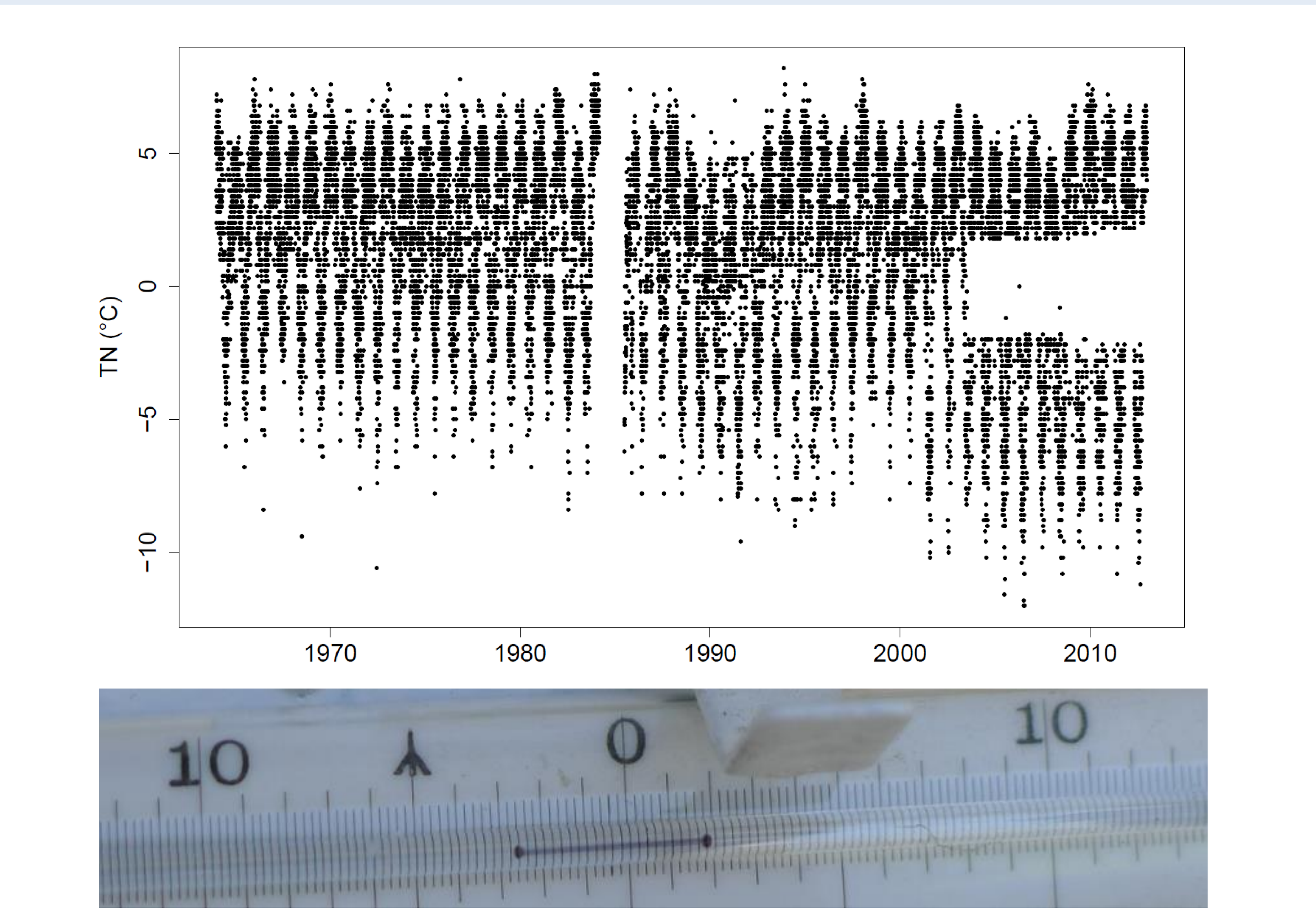

Fig. 3: Above: Since 2003, minimum temperature values between -2 and 2°C are missing at the station in Progreso, Peru. Below: The instrument used at the station. The value should be read at the right end of the rod.

B) 20mm precipitation cut

In the precipitation data of some stations, events of >20mm are partly or completely missing. In Aguirre, Bolivia, values >20mm are completely missing from 1989 onwards (Fig. 4, left). Investigating the rain gauge during station visits revealed the source of the error. The Hellmann rain gauge used in Bolivia are comprised of an inner container, placed in an outer container, which is toped by a cone (Fig. 4, right). The inner container collects 20mm of precipitation before overflowing. The observer measures the precipitation by sticking a scale in the inner container and reading the water mark on the scale. In precipitation events of >20mm, the observer needs to refill the inner container with the additional water from the outer container (up to 200mm) and cumulate the measurements. Apparently, some observers did not add the water in the outer container to the daily precipitation sum.

Missing high precipitation events in a time series is a severe problem, and the data cannot be used for analysis of precipitation sums. For studies based on wet days, however, the data is usable.

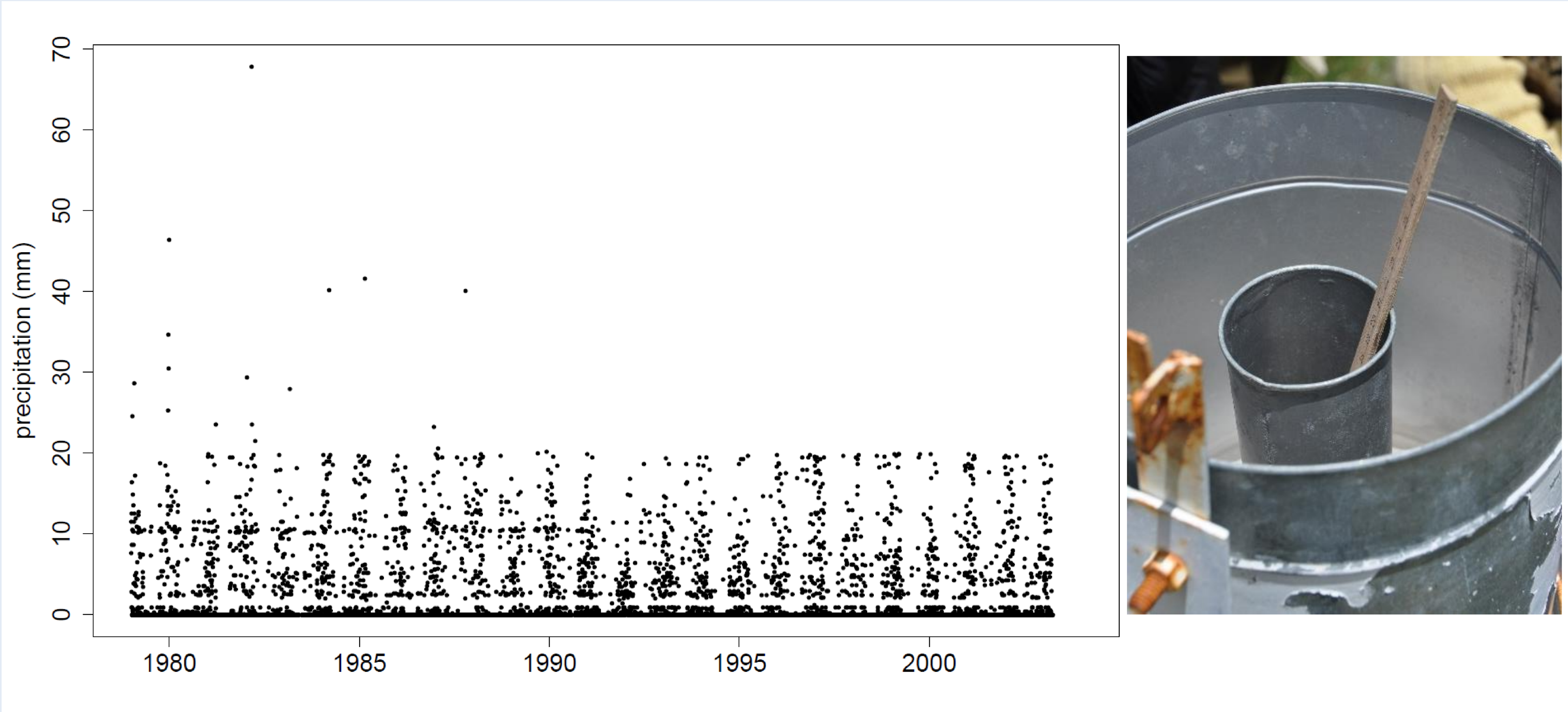

Fig. 4: Left: Only a few precipitation events >20mm are reported before 1989 in Aguirre, Bolivia. From 1989 onwards, measurements are completely truncated at 20mm. Right: Hellmann rain gauges are used in Bolivia.

6 Error detection

Various software exist to quality control climatological station data. These programs detect efficiently unreasonable values and outliers. However, some of the errors found in Bolivia and Peru are not detected by most standard tests.

Plotting and visually control the data is an effective way to detect a wide range of abnormalities. Many quality control software offer data plots for visual control.

Nevertheless, these visualizations are not suitable to clearly show all types of errors. This is demonstrated by the case of the missing values around 0°C in minimum temperatures in Progreso (Fig. 5 and 6). A suitable plot visualizes the error clearly (Fig. 3).

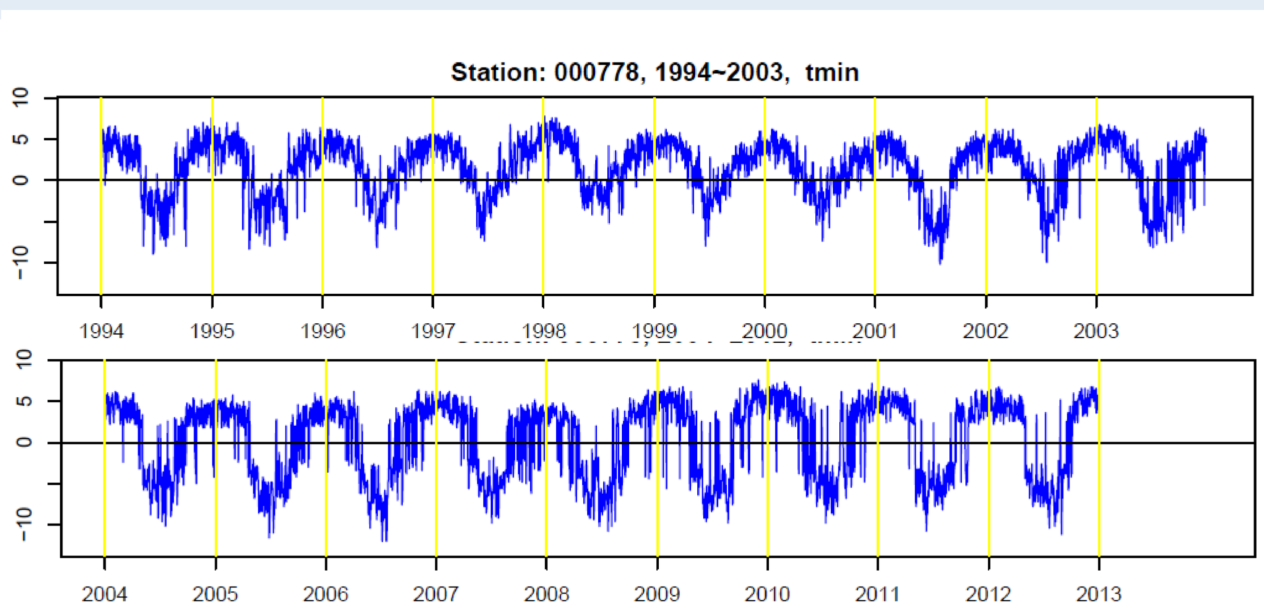

Fig. 5: RClimDex: The time series plot hides the range of missing values. The error would most likely remain undetected.

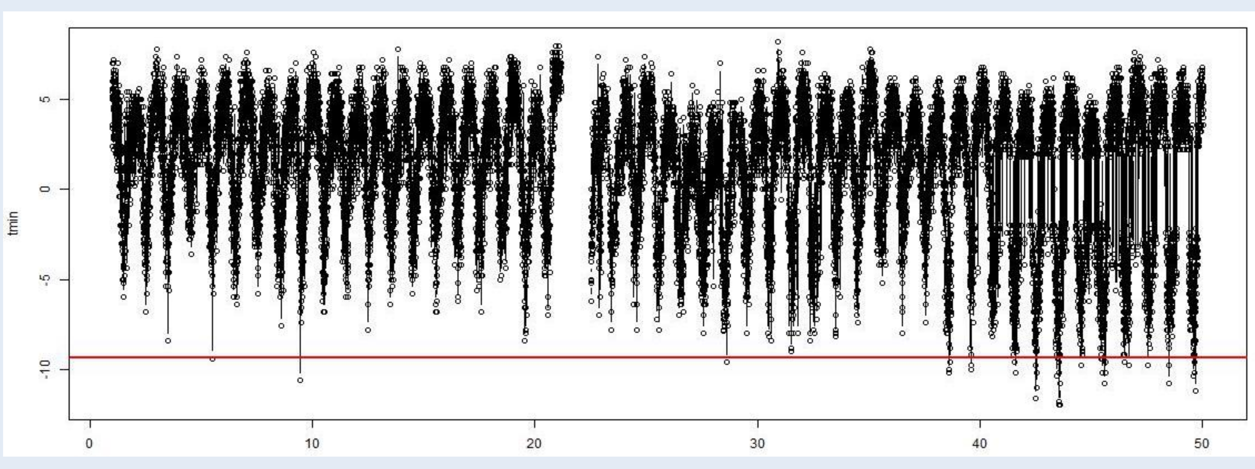

Fig. 6: RClimTool: The error is partly hidden by the lines between the dots. The red line indicates the outer bound of 3 times the standard deviation. From 2003 onwards, some more low values are regarded as suspicious.

7 Conclusions

- Quality control and assessment of Central Andean observation data is of great importance.
- Only applying a quality control software before analyzing the data may not be sufficient.
- A visual control of all stations using a scatter plot without lines connecting the data points is recommendable.
- Station visits are a very useful way to get station information if metadata is scarce. Furthermore, it offers the chance to detect sources of errors, what may enable error correction.
- Clear approaches on how to asses, control and improve data quality in countries such as Bolivia and Peru are missing. The authors are developing an Approach on how to use quality control software in combination with other methods to efficiently detect all common data problems.

References

[1] Buytaert, W., and B. De Bievre (2012) Water for cities: The impact of climate change and demographic growth in the tropical Andes. *Water Res. Res*. **48**, doi:10.1029/2011WR011755.

[2] Seiler, C., R.W.A. Hutjes, P. Kabat (2013) Climate variability and trends in Bolivia. *J. Appl. Met. Clim.* **52**, 130-146.

[3] Gubler et al. (2015) Homogenization of temperature and precipitation measurements in areas of sparse station networks applying HOMER. In review.

[4] Brugnara, Y., M. Brunetti, M. Maugeri, T. Nanni, C. Simolo (2011) High-resolution analysis of daily precipitation trends in the central Alps over the last century. *Int. J. Climatol*. **32**, 1406–1422.

[5] Vertacnik, G., M. Dolinar, R. Bertalanic, M. Klancar, D. Dvorsek and M. Nadbath (2015) Ensemble homogenization of Slovenian monthly air temperature seris. *Int. J. Climatol*. Doi: 10.1002/joc.4265.